EXPRESS MAILING CERTIFICATE

"EXPRESS MAIL" Mailing Label No.: ELOSY 649384 OS

Date of Deposit: Quart & 7 3001

I hereby certify that this paper or fee is being deposited

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Box Patent Application,

Washington, D.C. 20231 Typed or printed name of person signing this certificate:

Signed: Candiso morore

24024
PATENT TRADEMARK OFFICE

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:)	Examiner: Richard Hutson
Gorski et al.)	
)	Art Unit: 1652
Serial No.: 09/078,465)	
)	
Filed: May 14, 1998)	
	`	

For: HOMEOBOX GENE) Docket No.: 22311/04015

Assistant Commissioner of Patents

Washington, D.C. 20231

STATEMENT REGARDING COMPUTER READABLE FORM OF SEQUENCE LISTING

Dear Sir:

The computer readable form in the above-described continuation application is identical with the last-filed computer readable form submitted with parent Application No. 09/078,465. In accordance with 37 C.F.R. 1.821(e), please use the last-filed computer readable form of the sequence listing, which was filed in the parent Application on September 23, 1997, as the computer readable form of the sequence listing in the instant application. It is understood that the Patent and Trademark Office will make the necessary change in Application number and filing date for the computer readable form that will be used in the instant application.

A paper copy of the Sequence Listing is included in the specification of the instant, continuation application.

Respectfully submitted,

Dated: Quegrest 27, 2001

Pamela A. Docherty, Reg. No. 40,591

(216) 622-8416

SUBSTITUTE

SEQUENCE LISTING

- (1) GENERAL INFORMATION:
 - (i) APPLICANT: Gorski, David H. Walsh, Kenneth
 - (ii) TITLE OF INVENTION: Growth Arrest Homeobox Gene
 - (iii) NUMBER OF SEQUENCES: 19
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Calfee, Halter, and Griswold
 - (B) STREET: 800 Superior Avenue
 - (C) CITY: Cleveland
 - (D) STATE: Ohio
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 44114-2688
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
 - (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Golrick, Mary E.
 - (B) REGISTRATION NUMBER: 34829
 - (C) REFERENCE/DOCKET NUMBER: 22311/00114
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (216) 622-8200
 - (B) TELEFAX: (216) 241-0816
 - (C) TELEX: 980499
- (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2244 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 197..1108

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GTCAAGTGTT TATACGTGCA GGAGACTGGC CGCTCGGCTC AGGACTGGGA TTAGCGGGCT	60
CTGCTCAAAC CCGCGCGCT TTTACATTAG GAGTGAGTGG GGGAGAGTCC TAGGATTTCT	120
AGTGAAAAGT GACAGCGCTT GGTGGACTTT GGGACCTTCG TGAAGTCTTC TGCTTGGAAG	180
CTGAGACTTG CATGCC ATG GAA CAC CCC CTC TTT GGC TGC CTG CGC AGC Met Glu His Pro Leu Phe Gly Cys Leu Arg Ser 1 5 10	229
CCC CAC GCC ACA GCG CAA GGC TTG CAC CCC TTC TCG CAG TCT TCT CTG Pro His Ala Thr Ala Gln Gly Leu His Pro Phe Ser Gln Ser Ser Leu 15 20 25	<u>.</u> 277
GCC CTC CAT GGA AGA TCT GAC CAC ATG TCC TAC CCC GAA CTC TCC ACA Ala Leu His Gly Arg Ser Asp His Met Ser Tyr Pro Glu Leu Ser Thr 30 35 40	325
TCT TCC TCG TCT TGC ATA ATC GCG GGA TAC CCC AAT GAG GAG GGC ATG Ser Ser Ser Cys Ile Ile Ala Gly Tyr Pro Asn Glu Glu Gly Met 45 50 55	373
TTT GCC AGC CAG CAT CAC AGG GGG CAC CAC CAC CAC CAC CAC CAC	421
CAT CAC CAC CAG CAG CAG CAG CAC CAG GCT CTG CAA AGC AAC TGG His His His Gln Gln Gln His Gln Ala Leu Gln Ser Asn Trp 80 85 90	469
CAC CTC CCG CAG ATG TCC TCC CCG CCA AGC GCG GCC CGG CAC AGC CTT His Leu Pro Gln Met Ser Ser Pro Pro Ser Ala Ala Arg His Ser Leu 95 100 105	517
TGC CTG CAG CCT GAT TCC GGA GGG CCC CCG GAG CTG GGG AGC AGC CCT Cys Leu Gln Pro Asp Ser Gly Gly Pro Pro Glu Leu Gly Ser Ser Pro 110 115 120	565
CCG GTC CTG TGC TCC AAC TCT TCT AGC CTG GGC TCC AGC ACC CCG ACC Pro Val Leu Cys Ser Asn Ser Ser Ser Leu Gly Ser Ser Thr Pro Thr 125 130 135	613
GGA GCC GCG TGC GCA CCA AGG GAT TAT GGC CGT CAA GCG CTG TCA CCC	661

Gly Ala Ala Cys Ala Pro Arg Asp Tyr Gly Arg Gln Ala Leu Ser Pro

140					145					150)				155	
					Arg					Arg				AGT Ser 170	Ser	709
				Gly					Glu						AGG Arg	757
														GAG ı Glu	GCA Ala	805
														GAG Glu		853
											Lys			TTC Phe	CAG Gln 235	901
														GGA Gly 250	Ala	949
														CTT Leu		997
														ACA Thr		1045
Asp		Leu												TCT Ser		1093
CAC His 300				TGAT	'ACAT	'AC A	.GAGA	CCAG	C TC	CGTT	CTCA	GGA	AAG	CACC		1145
ATTG	TGAT	GG C	TAAA:	CTCA	rc cc	:AAAC	ATCG	TTI	'ACA'I	GGC	AGAT	'GACT	GT G	GCAG	TGTTG	1205
CTTA	TATA	T AA	TAAA	\CGCA	'G GC	ATCI	'CAAG	TCI	GTTI	CTC	ATGA	TTGA	TA G	AAGG	TTTAC	1265
ACTA	AGTG	CC T	CTTA	\TTGA	A GA	TGCT	TCCA	CAG	TGAA	TTA	GGAG	AAAG	TG A	ACAT	ATCTA	1325
AATA	TACT	TG T	TCCI	TATA.	T GA	CAGA	GAGG	GAG	ATGA	ATG	TTTG	CTTT	GG C	TTGC	ACTGA	1385
AAAT	TAAA	ATT G	CTAC	CAAG	A GC	AAAC	TCGG	TAA	GACA	TTT	TGAC	TCAA	GT I	GTCT	CCAGA	1445
GTGA	AGAT	GT T	'ATAC	TAAA	'G CT	'TTGA	ACAT	TCC	AGTI	GTA	CCAG	GTCA	TG I	GTGT	GACAC	1505

TGGGCAGGTA	TTTGCTTTTG	CTTGCACTGA	AACTTAAACT	GCTATCAAGT	TAACCCATGA	1569
AATAGTTTAT	CTTGAACAGC	CACAGTGCCT	GAAATCACCA	AGTGGATATA	AAATGAACTG	1625
AAATTCTGTA	TATATTACTC	CTAAGTCATT	TTCCTGTCTT	CACTAATTTT	AGCAAATGCA	1685
TTCATATTAG	CTGATGAAAA	TAGGCTTTCC	CGTGGACAAA	TGCAGCCAGC	TTCTTGTATT	1745
TTTATACATT	TTTTTGTCAG	TCAGAGACAT	CAGTATGTGC	TTACTTGTGT	TCAAGTAGAG	1805
GAAATGCAGT	AGAGTCTGAT	AGGACATATT	CTTGGTACCA	CAGACAAAAC	AAATCTTCTG	1865
TTGCATTGAC	TATCAACTGC	TGCAGATACA	TTAGAGAACA	CACCTAGCCC	CCCTCCAGCC	1925
TCCCTCTGTT	ATCGCTCGAA	GACATTAGCG	TCATAGGCAA	GTAGTTACCT	TGCCAAATGA	1985
GTCTTGTGTG	GCAGATGTCT	GATTTTGTAT	CTTTAAACTG	TTAATGGTAT	GTGTCTGCTT	2045
CAGTTAACAG	GGAAAAAGAT	TTCTTCCTCA	TTGTTTATGA	TACAAAACCC	AAGTGCCAAA	2105
CAAAGCTAGT	TCTTCAAGGG	ATAGATGAGA	AACTGAATGT	CTGACAAGTA	GACTCAGCGA	2165
AAATACATTA	TTTTTCAGAG	GCTGTGTATT	CATGCAGTAC	AAGTCCTTGT	ATTTTGTAAA	2225
AAAAAAAGTT	AAATAAATG					2244

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 303 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu His Pro Leu Phe Gly Cys Leu Arg Ser Pro His Ala Thr Ala 1 5 10 15

Gln Gly Leu His Pro Phe Ser Gln Ser Ser Leu Ala Leu His Gly Arg 20 25 30

Ser Asp His Met Ser Tyr Pro Glu Leu Ser Thr Ser Ser Ser Cys
35 40 45

Ile Ile Ala Gly Tyr Pro Asn Glu Glu Gly Met Phe Ala Ser Gln His 50 55 60

His Arg Gly His His His His His His His His His Gln
65 70 75 80

Gln Gln Gln His Gln Ala Leu Gln Ser Asn Trp His Leu Pro Gln Met

85 90 95^{*}

Ser Ser Pro Pro Ser Ala Ala Arg His Ser Leu Cys Leu Gln Pro Asp 100 105 110

Ser Gly Gly Pro Pro Glu Leu Gly Ser Ser Pro Pro Val Leu Cys Ser 115 120 125

Asn Ser Ser Ser Leu Gly Ser Ser Thr Pro Thr Gly Ala Ala Cys Ala 130 135 140

Pro Arg Asp Tyr Gly Arg Gln Ala Leu Ser Pro Ala Glu Val Glu Lys 145 150 155 160

Arg Ser Gly Ser Lys Arg Lys Ser Asp Ser Ser Asp Ser Gln Glu Gly 165 170 175

Asn Tyr Lys Ser Glu Val Asn Ser Lys Pro Arg Arg Glu Arg Thr Ala 180 185 190

Phe Thr Lys Glu Gln Ile Arg Glu Leu Glu Ala Glu Phe Ala His His 195 200 205

Asn Tyr Leu Thr Arg Leu Arg Arg Tyr Glu Ile Ala Val Asn Leu Asp 210 215 220

Leu Thr Glu Arg Gln Val Lys Val Trp Phe Gln Asn Arg Arg Met Lys 225 230 235 240

Trp Lys Arg Val Lys Gly Gly Gln Gln Gly Ala Ala Arg Glu Lys 255

Glu Leu Val Asn Val Lys Lys Gly Thr Leu Leu Pro Ser Glu Leu Ser 260 265 270

Gly Ile Gly Ala Ala Thr Leu Gln Gln Thr Gly Asp Ser Leu Ala Asn 275 280 285

Asp Asp Ser Arg Asp Ser Asp His Ser Ser Glu His Ala His Leu 290 295 300

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 941 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

ı	i	٧ ٧	FE	יאי	יזיד	т	בי	
١	ı ı.	χJ		A	ıι	ıĸ	Ľ.	:

(A) NAME/KEY: CDS

(B) LOCATION: 33..941

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

		•	•	~													
	GTC'	TTCT.	ACC '	TGGA	ACCC	GA A	ACTT(GCAT	G CT		Glu				Phe	GGC Gly	53
٠						CAC His			Ala					Pro		TCC Ser	101
						CTC Leu		Gly					Met			CCC Pro	149
	GAG Glu 40	CTC Leu	TCT Ser	ACT Thr	TCT Ser	TCC Ser 45	TCA Ser	TCT Ser	TGC Cys	ATA Ile	ATC Ile 50	Ala	GGA Gly	TAC Tyr	CCC Pro	AAC Asn 55	197
The state of the s						GCC Ala										His	245
						CAC His										CAA Gln	293
	ACC Thr	AAC Asn	TGG Trp ,90	CAC His	CTC Leu	CCG Pro	CAG Gln	ATG Met 95	TCT Ser	TCC Ser	CCA Pro	CCG Pro	AGT Ser 100	GCG Ala	GCT Ala	CGG Arg	341
	CAT His	AGC Ser 105	CTC Leu	TGC Cys	CTC Leu	CAG Gln	CCC Pro 110	GAC Asp	TCT Ser	GGA Gly	GGG Gly	CCC Pro 115	CCA Pro	GAG Glu	TTG Leu	GGG Gly	389
•						CTG Leu 125										AGC Ser 135	437
<i>'</i>						GCG Ala											485
	CTG Leu	TCA Ser	CCT Pro	GCG Ala 155	GAG Glu	GCG Ala	GAG Glu	AAG Lys	CGA Arg 160	AGC Ser	GGC Gly	GGC Gly	AAG Lys	AGG Arg 165	AAA Lys	AGC Ser	533

	GAC Asp	AGC Ser	TCA Ser 170	GAC Asp	TCC Ser	CAG Gln	GAA Glu	GGA Gly 175	Asn	TAC Tyr	AAG Lys	TCA Ser	GAA (Glu 180	GTC I Val	AAC <i>I</i> Asn	AGC Ser	5	581
r '													CAA A				6	529
												Thr	AGA (Arg				6	577
													CAG (Gln				7	725
ana,													AAG (Lys				7	773
													GTG A Val 260				8	321
													GCC A Ala				8	869
i Ö													GAC A				9	17
	AGC Ser							TGA									9	41

(2) INFORMATION FOR SEQ ID NO:4:

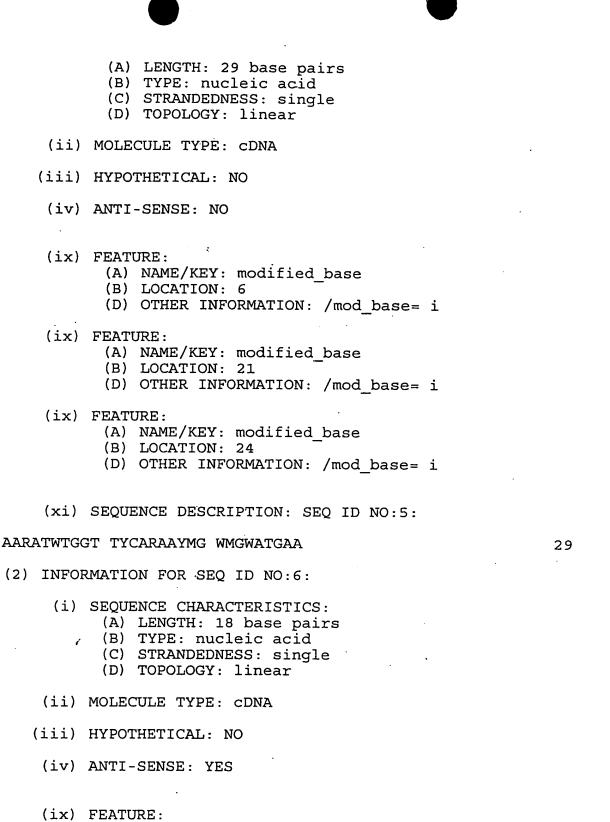
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 302 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Glu His Pro Leu Phe Gly Cys Leu Arg Ser Pro His Ala Thr Ala 1 5 10 15

Gln Gly Leu His Pro Phe Ser Gln Ser Ser Leu Ala Leu His Gly Arg

Ser Asp His Met Ser Tyr Pro Glu Leu Ser Thr Ser Ser Ser Cys Ile Ile Ala Gly Tyr Pro Asn Glu Glu Asp Met Phe Ala Ser Gln His His Arg Gly His His His His His His His His His Gln Gln Gln Gln His Gln Ala Leu Gln Thr Asn Trp His Leu Pro Gln Met Ser Ser Pro Pro Ser Ala Ala Arg His Ser Leu Cys Leu Gln Pro Asp Ser Gly Gly Pro Pro Glu Leu Gly Ser Ser Pro Pro Val Leu Cys Ser Asn 115 120 Ser Ser Ser Leu Gly Ser Ser Thr Pro Thr Gly Ala Ala Cys Ala Pro 135 Gly Asp Tyr Gly Arg Gln Ala Leu Ser Pro Ala Glu Ala Glu Lys Arg 150 Ser Gly Gly Lys Arg Lys Ser Asp Ser Ser Asp Ser Gln Glu Gly Asn Tyr Lys Ser Glu-Val Asn Ser Lys (Pro Arg Lys Glu Arg Thr Ala Phe 180 185 Thr Lys Glu Gln Ile Arg Glu Leu Glu Ala Glu Phe Ala His His Asn 200 205 Tyr Leu Thr Arg Leu Arg Arg Tyr Glu Ile Ala Val Asn Leu Asp Leu 210 Thr Glu Arg Gln Val Lys Val Trp Phe Gln Asn Arg Arg Met Lys Trp 230 235 . Lys Arg Val Lys Gly Gln Gln Gly Ala Ala Ala Arg Glu Lys Glu 250 Leu Val Asn Val Lys Lys Gly Thr Leu Leu Pro Ser Glu Leu Ser Gly 260 265 Ile Gly Ala Ala Thr Leu Gln Gln Thr Gly Asp Ser Ile Ala Asn Glu 275 285 Asp Ser His Asp Ser Asp His Ser Ser Glu His Ala His Leu 295 300

- (2) INFORMATION FOR SEQ ID NO:5:
 - (i) SEQUENCE CHARACTERISTICS:



(A) NAME/KEY: modified base

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

(D) OTHER INFORMATION: /mod_base= i

(B) LOCATION: 4



- (2) INFORMATION FOR SEQ ID NO:7:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCGCGCAGAT CTCACTGAAA GACAGGTAAA

30

- (2) INFORMATION FOR SEQ ID NO:8:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: YES
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTTACCTGTC TTTCAGTGAG

- (2) INFORMATION FOR SEQ ID NO:9:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO

(1V) ANTI-SENSE: YES	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	
GCGCGCAGAT CTAGATTCAC TGCTATCTCG TA	32
(2) INFORMATION FOR SEQ ID NO:10:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: YES	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:	
GCGCGTGCCC CCTCTGATGC TGGCTGGCAA ACATGT	36
(2) INFORMATION FOR SEQ ID NO:11:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 32 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(iii) HYPOTHETICAL: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
GCGCGCTCTT GAAGGGCGAG AGAGGATTGG GA	32
(2) INFORMATION FOR SEQ ID NO:12:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 38 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	

(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
	•
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
CTGGTTCGGC CCACCTCTGA AGGTTCCAGA ATCGATAG	38
(2) INFORMATION FOR SEQ ID NO:13:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
GGAGACTTCC AAGGTCTTAG CTATCACTTA AGCAC	35
(2) INFORMATION FOR SEQ ID NO:14:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(iii) HYPOTHETICAL: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
GCGCGCGTCG ACGAACACCC CCTCTTTGGC	30
(2) INFORMATION FOR SEQ ID NO:15:	
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 30 base pairs(B) TYPE: nucleic acid	

(ii) MOLECULE TYPE: cDNA	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:	
GCGCGCAAGC TTTCATAAGT GTGCGTGCTC	30
(2) INFORMATION FOR SEQ ID NO:16:	
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 25 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	•
CCCGCGCGCC TTTTACATTA GGAGT	25
(2) INFORMATION FOR SEQ ID NO:17:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	•
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
îÖ (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:	
GCTGGCAAAC ATGCCCTCCT CATTG	25
(2) INFORMATION FOR SEQ ID NO:18:	

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i) -	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(iii)	HYPOTHETICAL: NO	
(iy)	ANTI-SENSE: NO	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:18:	
TGATGGCA	TG GACTGTGGTC ATGA	24
(2) INFO	RMATION FOR SEQ ID NO:19:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(iii)	HYPOTHETICAL: NO	
(iv)	ANTI-SENSE: NO	
,		
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:19:	
TGATGGCA	G GACTGTGGTC ATGA	24

65

SEQUENCE LISTING

, ,5	(1) GENERAL INFORMATION:
,3	(i) APPLICANT: Gorski, David H. Walsh, Kenneth
	(ii) TITLE OF INVENTION: Growth Arrest Homeobox Gene
10	(iii) NUMBER OF SEQUENCES: 4
15	(iv) CORRESPONDENCE ADDRESS: (A) ADDRESSEE: Calfee, Halter, and Griswold (B) STREET: 800 Superior Avenue (C) CITY: Cleveland
20	(D) STATE: Ohio (E) COUNTRY: U.S.A. (F) ZIP: 44114-2688
	 (v) COMPUTER READABLE FORM: (A) MEDIUM TYPE: Floppy disk (B) COMPUTER: IBM PC compatible (C) OPERATING SYSTEM: PC-DOS/MS-DOS (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
30 30	<pre>(vi) CURRENT APPLICATION DATA: (A) APPLICATION NUMBER: (B) FILING DATE: (C) CLASSIFICATION:</pre>
135 135	(viii) ATTORNEY/AGENT INFORMATION: (A) NAME: Golrick, Mary E. (B) REGISTRATION NUMBER: 34829 (C) REFERENCE/DOCKET NUMBER: 22311/00114
40	(ix) TELECOMMUNICATION INFORMATION: (A) TELEPHONE: (216) 622-8200 (B) TELEFAX: (216) 241-0816 (C) TELEX: 980499
	(2) INFORMATION FOR SEQ ID NO:1:
45	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 2244 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: both (D) TOPOLOGY: linear
//	(ii) MOLECULE TYPE: CDNA
	(iii) HYPOTHETICAL: NO
55	(iv) ANTI-SENSE: NO
60	(ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 1971108
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GTCAAGTGTT TATACGTGCA GGAGACTGGC CGCTCGGCTC AGGACTGGGA TTAGCGGGCT 60

and provide the

		CTGC 120	CTCA	AAC (CCGC	CGG	CT T	TTAC	ATTAC	GA(GTGA	GTGG	GGG:	AGAG	TCC	TAGG	ATTTCT
<u>,</u>		AGTO	AAA	AGT (GACAC	GCGCT	TT G	GTGG	ACTTT	GGC	GACC'	TTCG	TGA	AGTC'	TTC	TGCT	TGGAAG
		CTGA 229	\GAC1	TTG (CATGO	CC AT	rg gi	AA CI	AC C	CC C	rc T	TT G	GC T	GC C	TG C	GC A	GC
10						Me	et Gl	lu Hi	s Pr	o Le	u Ph 5	e Gl	у Су	s Le	u Ar 1	g Se .0	r
		CCC 277	CAC	GCC	ACA	GCG	CAA	GGC	TTG	CAC	CCC	TTC	TCG	CAG	TCT	TCT	CTG
15	• •		His	Ala	Thr 15	Ala	Gln	Gly	Leu	His 20	Pro	Phe	Ser	Gln	Ser 25	Ser	Leu
	٠.	GCC 325	CTC	CAT	GGA	AGA	TCT	GAC	CAC	ATG	TCC	TAC	CCC	GAA	CTC	TCC	ACA
2 0			Leu	His 30	Gly	Arg	Ser	.Asp	His 35	Met	Ser	Tyr	Pro	Glu 40	Leu	Ser	Thr
o O		TCT. 373	TCC	TCG	TCT	TGC	ATA	ATC	GCG	GGA	TAC	ccc	, AAT	GAG	GAG	GGC	ATG
2 5			Ser 45	Ser	Ser	Суз	Ile	Ile 50	Ala	Gly	Tyr	Pro	Asn 55	Glu	Glu	Gly	Met
) (1)		TTT 421	GCC	AGC	CAG	CAT	CAC	AGG	GGG	CAC							
30	•		Ala	Ser	Gln	His	His 65	Arg	Gly	His	His	His 70	His	His	His	His	His 75
: 0 m		CAT	CAC	CAC	CAC	CAG	CAG	CAG	CAG	CAC	CAG	GCT	CTG	CAA	AGC	AAC	TGG.
[] [35			His	His	His	Gln 80	Gln	Gln	Gln	His	Gln 85	Ala	Leu	Gln	Ser	Asn 90	Trp
7 0		CAC 517	CTC	CCG	CAG	ATG	TCC	TCC	CCG	CCA	AGC	GCG	GCC	CGG	CAC	AGC	CTT
⊭≟ 40		His	Leu	Pro	Gln 95	Met	Ser	Ser	Pro	Pro 100	Ser	Ala	Ala	Arg	His 105	Ser	Leu
		TGC 565	CTG	CAG	CCT	GAT	TCC	GGA	GGG	CCC	CCG	GAG	CTG	GGG	AGC	AGC	CCT
45		CÃa	Leu	Gln 110	Pro	yab	Ser	Gly	Gly 115	Pro	Pro	Glu	Leu	Gly 120	Ser	Ser	Pro
		CCG 613	GTC	CIG	TGC	TCC	AAC	TCT	TCT	AGC	CTG	GGC	TCC	AGC	ACC	CC	ACC
50		Pro	Val 125	Leu	Сув	Ser	Asn	Ser 130	Ser	Ser	Leu	Gly	Ser 135	Ser	Thr	Pro	Thr
		GGA 661	GCC	GCG	TĢC	GCA	CCA	AGG	GAT	TAT	GGC	CGT	CAA	GCG	CTG	TCF	CCC
55		_	Ala	Ala	Cys	Ala	Pro 145	Arg	Asp	Tyr	Gly	Arg 150	Gln	Ala	Leu	Ser	Pro 155
		GCA 709	GAA	GTG	GAG	AAG	AGA	AGT	GGC	AGC	ÄAA	A AGA	AAA	AĢC	GAC	AG1	TCA
60			Glu	Val	Glu	Lys 160	Arg	Ser	Gly	Ser	Lys 165		ГÀа	Ser	Asp	Ser 170	Ser
		GAT 757	TCC	CAG	GAA	GGA	. AAT	TAC	AAG	TCA	GA#	A GTG	AAC	AGO	: AA	A CC	r Agg
. 65			Ser	Gln	Glu 175	Gly	Asn	Tyr	Lys	Ser 180	Glu	Val	Asn	Ser	Lys 185	Pro	Arg

AGG GAA AGA ACA GCT TTC ACC AAA GAG CAA ATC AGA GAA CTT GAG GCA

22311/00114

	805 Arg Glu Arg Thr Ala Phe Thr Lys Glu Gln Ile Arg Glu Leu Glu Ala 190 195 200
5	GAG TTC GCC CAT CAT AAC TAT CTG ACC AGA CTG AGA AGA TAT GAG ATA 853
	Glu Phe Ala His His Asn Tyr Leu Thr Arg Leu Arg Arg Tyr Glu Ile 205 210 215
10 .	GCG GTG AAC CTA GAC CTC ACT GAA AGA CAG GTG AAA GTG TGG TTC CAG 901
	Ala Val Asn Leu Asp Leu Thr Glu Arg Gln Val Lys Val Trp Phe Gln 220 235
15	AAC AGG AGA ATG AAG TGG AAG CGG GTC AAG GGG GGA CAA CAA GGA GCT 949
	Asn Arg Arg Met Lys Trp Lys Arg Val Lys Gly Gln Gln Gly Ala 240 245 250
20	GCA GCC CGA GAA AAG GAA CTG GTG AAT GTG AAA AAG GGA ACA CTT CTT
	Ala Ala Arg Glu Lys Glu Leu Val Asn Val Lys Lys Gly Thr Leu Leu 255 260 265
9 4	CCA TCA GAG CTG TCA GGA ATT GGT GCA GCC ACC CTC CAG CAG ACA GGG
<u>دُ</u> 1	Pro Ser Glu Leu Ser Gly Ile Gly Ala Ala Thr Leu Gln Gln Thr Gly 270 275 280
3.0 LLJ	GAC TCA CTA GCA AAT GAC GAC AGT CGC GAT AGT GAC CAC AGC TCT GAG
e 	Asp Ser Leu Ala Asn Asp Asp Ser Arg Asp Ser Asp His Ser Ser Glu 285 290 295
35	CAC GCA CAC TTA TGATACATAC AGAGACCAGC TCCGTTCTCA GGAAAGCACC
35 U	CAC GCA CAC TTA TGATACATAC AGAGACCAGC TCCGTTCTCA GGAAAGCACC 1145 His Ala His Leu 300
35	1145 His Ala His Leu
3階 引 マ 46	1145 His Ala His Leu 300 ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG
35	His Ala His Leu 300 ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG 1205 CTTAATATAA TTAAACGCAG GCATCTCAAG TCTGTTTCTC ATGATTGATA GAAGGTTTAC
3階 引 マ 46	His Ala His Leu 300 ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG 1205 CTTAATATAA TTAAACGCAG GCATCTCAAG TCTGTTTCTC ATGATTGATA GAAGGTTTAC 1265 ACTAAGTGCC TCTTATTGAA GATGCTTCCA CAGTGAAATT GGAGAAAGTG AACATATCTA
35 10 √ 0 40 45	His Ala His Leu ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG 1205 CTTAATATAA TTAAACGCAG GCATCTCAAG TCTGTTTCTC ATGATTGATA GAAGGTTTAC 1265 ACTAAGTGCC TCTTATTGAA GATGCTTCCA CAGTGAAATT GGAGAAAGTG AACATATCTA 1325 AATATACTTG TTCCTTATAT GACAGAGAGG GAGATGAATG TTTGCTTTGG CTTGCACTGA
35 10 √ 0 40 45	His Ala His Leu ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG 1205 CTTAATATAA TTAAACGCAG GCATCTCAAG TCTGTTTCTC ATGATTGATA GAAGGTTTAC 1265 ACTAAGTGCC TCTTATTGAA GATGCTTCCA CAGTGAAATT GGAGAAAGTG AACATATCTA 1325 AATATACTTG TTCCTTATAT GACAGAGAGG GAGATGAATG TTTGCTTTGG CTTGCACTGA AAAATTAAATT GCTACCAAGA GCAAACTCGG TAAGACATTT TGACTCAAGT TGTCTCCAGA
35 10 40 45	His Ala His Leu ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG 1205 CTTAATATAA TTAAACGCAG GCATCTCAAG TCTGTTTCTC ATGATTGATA GAAGGTTTAC 1265 ACTAAGTGCC TCTTATTGAA GATGCTTCCA CAGTGAAATT GGAGAAAGTG AACATATCTA 1325 AAATATACTTG TTCCTTATAT GACAGAGAGG GAGATGAATG TTTGCTTTGG CTTGCACTGA 1385 AAATTAAAATT GCTACCAAGA GCAAACTCGG TAAGACATTT TGACTCAAGT TGTCTCCAGA GTGAAAGATGT TATAGAAATG CTTTGAACAT TCCAGTTGTA CCAGGTCATG TGTGTGACAC
45 45 50	His Ala His Leu ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG 1205 CTTAATATAA TTAAACGCAG GCATCTCAAG TCTGTTTCTC ATGATTGATA GAAGGTTTAC 1265 ACTAAGTGCC TCTTATTGAA GATGCTTCCA CAGTGAAATT GGAGAAAGTG AACATATCTA 1325 AATATACTTG TTCCTTATAT GACAGAGAGG GAGATGAATG TTTGCTTTGG CTTGCACTGA 1385 AAATTAAATT GCTACCAAGA GCAAACTCGG TAAGACATTT TGACTCAAGT TGTCTCCAGA GTGAAGATGT TATAGAAATG CTTTGAACAT TCCAGTTGTA CCAGGTCATG TGTGTGACAC TGGGCAGGTA TTTGCTTTTG CTTGCACTGA AACTTAAACT GCTATCAAGT TAACCCATGA
45 45 50	His Ala His Leu 300 ATTGTGATGG CAAATCTCAC CCAAACATCG TTTACATGGC AGATGACTGT GGCAGTGTTG 1205 CTTAATATAA TTAAACGCAG GCATCTCAAG TCTGTTTCTC ATGATTGATA GAAGGTTTAC 1265 ACTAAGTGCC TCTTATTGAA GATGCTTCCA CAGTGAAATT GGAGAAAGTG AACATATCTA 1325 AATATACTTG TTCCTTATAT GACAGAGAGG GAGATGAATG TTTGCTTTGG CTTGCACTGA 1385 AAAATTAAAATT GCTACCAAGA GCAAACTCGG TAAGACATTT TGACTCAAGT TGTCTCCAGA 1445 GTGAAGATGT TATAGAAATG CTTTGAACAT TCCAGTTGTA CCAGGTCATG TGTGTGACAC 1505 TGGGCAGGTA TTTGCTTTTG CTTGCACTGA AACTTAAACT GCTATCAAGT TAACCCATGA AAATGATTAT CTTGAACAGC CACAGTGCCT GAAATCACCA AGTGGATATA AAATGAACTG

22311/00114

1745

[J] =

N

45

50.

55

60

65

TTTATACATT TTTTTGTCAG TCAGAGACAT CAGTATGTGC TTACTTGTGT TCAAGTAGAG 5 GAAATGCAGT AGAGTCTGAT AGGACATATT CTTGGTACCA CAGACAAAAC AAATCTTCTG 1865 TTGCATTGAC TATCAACTGC TGCAGATACA TTAGAGAACA CACCTAGCCC CCCTCCAGCC 10 1925 TCCCTCTGTT ATCGCTCGAA GACATTAGCG TCATAGGCAA GTAGTTACCT TGCCAAATGA 15 GTCTTGTGTG GCAGATGTCT GATTTTGTAT CTTTAAACTG TTAATGGTAT GTGTCTGCTT CAGTTAACAG GGAAAAAGAT TTCTTCCTCA TTGTTTATGA TACAAAACCC AAGTGCCAAA .20 CAAAGCTAGT TCTTCAAGGG ATAGATGAGA AACTGAATGT CTGACAAGTA GACTCAGCGA O AAATACATTA TTTTTCAGAG GCTGTGTATT CATGCAGTAC AAGTCCTTGT ATTTTGTAAA 25, 400 2225 AAAAAAGTT AAATAAATG 2244 30

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 303 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu His Pro Leu Phe Gly Cys Leu Arg Ser Pro His Ala Thr Ala 1 5 10 15

Gln Gly Leu His Pro Phe Ser Gln Ser Ser Leu Ala Leu His Gly Arg
20 25 30

Ser Asp His Met Ser Tyr Pro Glu Leu Ser Thr Ser Ser Ser Cys
35 40 45

Ile Ile Ala Gly Tyr Pro Asn Glu Glu Gly Met Phe Ala Ser Gln His 50 55 60

Gln Gln Gln His Gln Ala Leu Gln Ser Asn Trp His Leu Pro Gln Met 85 90 95

Ser Ser Pro Pro Ser Ala Ala Arg His Ser Leu Cys Leu Gln Pro Asp 100 105 110

Ser Gly Gly Pro Pro Glu Leu Gly Ser Ser Pro Pro Val Leu Cys Ser 115 120 125

Asn Ser Ser Ser Leu Gly Ser Ser Thr Pro Thr Gly Ala Ala Cys Ala 130 135 140

. ,		
		* • • •
	22311/00114 37	
	Pro Arg Asp Tyr Gly Arg Gln Ala Leu Ser Pro Ala Glu Val Glu Lys 145 150 155 160	
5	Arg Ser Gly Ser Lys Arg Lys Ser Asp Ser Ser Asp Ser Gln Glu Gly 165 170 175	
	Asn Tyr Lys Ser Glu Val Asn Ser Lys Pro Arg Arg Glu Arg Thr Ala 180 185 190	
10	Phe Thr Lys Glu Gln Ile Arg Glu Leu Glu Ala Glu Phe Ala His His 195 200 205	
15	Asn Tyr Leu Thr Arg Leu Arg Arg Tyr Glu Ile Ala Val Asn Leu Asp 210 215 220	
	Leu Thr Glu Arg Gln Val Lys Val Trp Phe Gln Asn Arg Arg Met Lys 235 230 235 240	
20	Trp Lys Arg Val Lys Gly Gly Gln Gln Gly Ala Ala Arg Glu Lys 245 250 255	
	Glu Leu Val Asn Val Lys Lys Gly Thr Leu Leu Pro Ser Glu Leu Ser 260 265 270	:
25 25 70 0	Gly Ile Gly Ala Ala Thr Leu Gln Gln Thr Gly Asp Ser Leu Ala Asn 275 280 285	
() () ()	Asp Asp Ser Arg Asp Ser Asp His Ser Ser Glu His Ala His Leu 290 295 300	
y	(2) INFORMATION FOR SEQ ID NO:3:	
3, E 7 I	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 941 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: both (D) TOPOLOGY: linear 	
40	(ii) MOLECULE TYPE: cDNA	
	(iii) HYPOTHETICAL: NO	
45	(iv) ANTI-SENSE: NO	
50·	(ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 33941	•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:	
55	GTCTTCTACC TGGAACCCGA AACTTGCATG CT ATG GAA CAC CCG CTC TTT GGC	
	Met Glu His Pro Leu Phe Gly 1 5	
60	TGC CTG CGC AGC CCT CAC GCC ACG GCG CAA GGC TTG CAC CCG TTC TCC 101 Cys Leu Arg Ser Pro His Ala Thr Ala Gln Gly Leu His Pro Phe Ser	
	10 15 20 CAA TCC TCT CTC GCC CTC CAT GGA AGA TCT GAC CAT ATG TCT TAC CCC	
65	149 Gln Ser Ser Leu Ala Leu His Gly Arg Ser Asp His Met Ser Tyr Pro 25 30 35	

	GNG (מייירי . יו	المل	a Cm	ጥረጥ	mcc	TC N	mca.	mc c	3003	3.00	000				
	197															CAAC
5	Glu I 40	Leu S	Ser	Thr	Ser	Ser 45	Ser	Ser	Сув	Ile	Ile 50	Ala	Gly	Tyr	Pro	Asn 55
	GAA (GAG G	BAC	ATG	TTT	GCC	AGC	CAG	CAT	CAC	AGG	GGG	CAC	CAC	CAC	CAC
	Glu (Glu A	/ab	Met	Phe 60	Ala	Ser	Gln	His	His 65	Arg	Gly	His	His	His 70	His
10	CAC	CAC C	LAC	CAT	CAC	CAC	CAT	CAG	CAG	CAG	CAG	CAC	CAC	COT		
	293															
4.5	His E	ire u	ITS	75	UIS	HIS	ura	GIN	80	GIN	GIN	HIB	GIn	Ala 85	Leu	Gln
15	ACC #	AAC I	GG	CAC	CTC	CCG	CAG	ATG	TCT	TCC	CCA	CCG	AGT	GCG	GCI	' CGG
	Thr A	Asn T	rp 90	His	Leu	Pro	Gln	Met 95	Ser	Ser	Pro	Pro	Ser 100	Ala	Ala	Arg
20	CAT A	AGC C	TC	TGC	CTC	CAG	ccċ	GAC	TCT	GGA	GGG	CCC	CCA	GAG	TTG	GGG
25 25	389 His S	er L	eu	Суз	Leu	Gln	Pro 110	Asp	Ser	Gly	Gly	Pro 115	Pro	Glu	Leu	Gly
25	AGC A		CG	CCC	GTC	CTG		TCC	AAC	TCT	TCC		TTG	GGC	TCC	: AGC
	437 Ser S															
J1 30	120					125				•	130			· · ¯.		135
	ACC 0	CG A	CT	GGG	GCC	GCG	TGC	GCG	CCG	GGG	GAC	TAC	GGC	CGC	CAG	GCA
: [] 35	Thr P	ro T	hr (Gly	Ala 140	Ala	Суз	Ala	Pro	Gly 145	Asp	Tyr	Gly	Arg	Gln 150	Ala
	CTG 1	CA C	CT	GCG	GAG	GCG	GAG	AAG	CGA	AGC	GGC	GGC	AAG	AGG	AAA	AGC
150 - 7 (Jo. 40)	Leu S	er P	ro :	Ala 155	Glu	Ala	Glu	Lys	Arg 160	Ser	Gly	Gly	Lys	Arg 165	Lys	Ser
. 42	GAC A	GC T	CA	GAC	TCC	CAG	GAA	GGA	AAT	TAC	AAG	TCA	GAA	GTC	AAC	AGC
45	Asp S	er S	er :	Asp	Ser	Gln	Glu	Gly 175	Asn	Tyr	Lys	Ser	Glu 180	Val	Asn	Ser .
45	AAA C	CC A	.GG	AAA	GAA	AGG	ACA	GCA	TTT	ACC	AAA	GAG	CAA	ATC	AGA	. GAA
'	Lys P	ro A .85	rg :	Lys	Glu	Arg	Thr 190	Ala	Phe	Thr	Lys	Glu 195	Gln	Ile	Arg	Glu
50·	CTT G	AA G	CA	GAA	TTT	GCC	CAT	CAT	AAT	TAT	CTC	ACC	AGA	CTG	AGG	CGA
4	Leu G 200	lu A	la (Glu	Phe	Ala 205	His	His	Asn	Tyr	Leu 210	Thr	Arg	Leu	Arg	Arg 215
55	TAC G	GAG A	TA	GCA	GTG	AAT	CTG	GAT	CTC	ACT		AGA	CAG	GTA	. AAA	
	725 Tyr G	lu I	le i	Ala	Val 220	Asn .	Leu	Asp	Leu		Glu	Arg	Gln	Val		Val
60	me-s -							_		225					230	
	TGG 1															
65	Trp P	he G	ln :	Asn 235	Arg	Arg	Met	Lys	Trp 240	Lys	Arg	Val	Lys	Gly 245	Gly	Gln
	CAA G 821	GA G	CT	GCG	GCT	CGG	GAA	AAG	GAA	CTG	GTG	AAT	GTG	AAA	. AAG	GGA

22311/00114

65

Gin Gly Ala Ala Arg Glu Lys Glu Leu Val Asn Val Lys Lys Gly ACA CTT CTC CCA TCA GAG CTG TCG GGA ATT GGT GCA GCC ACC CTC CAG 5 Thr Leu Leu Pro Ser Glu Leu Ser Gly Ile Gly Ala Ala Thr Leu Gln CAA ACA GGG GAC TCT ATA GCA AAT GAA GAC AGT CAC GAC AGT GAC CAC 10 Gln Thr Gly Asp Ser Ile Ala Asn Glu Asp Ser His Asp Ser Asp His AGC TCA GAG CAC GCC CAC CTC TGA 15 941 Ser Ser Glu His Ala His Leu 300 20 (2) INFORMATION FOR SEQ ID NO:4: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 302 amino acids (B) TYPE: amino acid 2年00 (D) TOPOLOGY: linear (ii) MOLECULE TYPE: protein · (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4: Met Glu His Pro Leu Phe Gly Cys Leu Arg Ser Pro His Ala Thr Ala Gln Gly Leu His Pro Phe Ser Gln Ser Ser Leu Ala Leu His Gly Arg 25 Ser Asp His Met Ser Tyr Pro Glu Leu Ser Thr Ser Ser Ser Cys ٠, 40 Ile Ile Ala Gly Tyr Pro Asn Glu Glu Asp Met Phe Ala Ser Gln His 45 Gln Gln His Gln Ala Leu Gln Thr Asn Trp His Leu Pro Gln Met Ser Ser Pro Pro Ser Ala Ala Arg His Ser Leu Cys Leu Gln Pro Asp Ser 50 100 105 Gly Gly Pro Pro Glu Leu Gly Ser Ser Pro Pro Val Leu Cys Ser Asn 120 55 Ser Ser Ser Leu Gly Ser Ser Thr Pro Thr Gly Ala Ala Cys Ala Pro 130 Gly Asp Tyr Gly Arg Gln Ala Leu Ser Pro Ala Glu Ala Glu Lys Arg 150 60

Ser Gly Gly Lys Arg Lys Ser Asp Ser Ser Asp Ser Gln Glu Gly Asn

Tyr Lys Ser Glu Val Asn Ser Lys Pro Arg Lys Glu Arg Thr Ala Phe

Thr Lys Glu Gln Ile Arg Glu Leu Glu Ala Glu Phe Ala His His Asn

195

5	Tyr	Leu 210	Thr	Arg	Leu	Arg	Arg 215	Tyr	Glu	Ile	Ala	Val 220	Asn	Leu	Asp	Leu
<i>e</i> ∨	. Thr 225	-Glu	Arg	Gln	Val	Lys 230	Val	Trp	Phe	Gln	Asn 235	Arg	Arg	Met	Lys	Trp 240
10	Lys	Arg	Val	ГЛа	Gly 245	Gly	Gln	Gln	Gly	Ala 250	Ala	Ala	Arg	Glu	Lys 255	Glu
	Leu	Val	Asn	Val 260	ГÄЗ	Lys	Gly	Thr	Leu 265	Leu	Pro	Ser	Glu	Leu 270	Ser	Gly
15	Île	Gly	Ala 275	Ala	Thr	Leu	Gln	Gln 280	Thr	Gly	Asp	Ser	Ile 285	Ala	Asn	Glu
	Asp	Ser 290	His	Asp	Ser	Asp	His 295	Ser	Ser	Glu	His	Ala 300	His	Leu		